Published in final edited form as:
Anal Chem. 2004 March 1; 76(5): 1243. doi:10.1021/ac0351163.

Influence of Basic Residue Content on Fragment Ion Peak Intensities in Low-Energy Collision-Induced Dissociation Spectra of Peptides

David L. Tabb[†], **Yingying Huang**[‡], **Vicki H. Wysocki**[‡], and **John R. Yates III**[*,§]
SR11 Department of Cell Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037

Abstract

The primary utility of trypsin digestion in proteomics is that it cleaves proteins at predictable locations, but it is also notable for yielding peptides that terminate in basic arginine and lysine residues. Tryptic peptides fragment in ion trap tandem mass spectrometry to produce prominent C-terminal y series ions. Alternative proteolytic digests may produce peptides that do not follow these rules. In this study, we examine 2568 peptides generated through proteinase K digestion, a technique that produces a greater diversity of basic residue content in peptides. We show that the position of basic residues within peptides influences the peak intensities of b and y series ions; a basic residue near the N-terminus of a peptide can lead to prominent b series peaks rather than the intense y series peaks associated with tryptic peptides. The effects of presence and position for arginine, lysine, and histidine are explored separately and in combination. Arg shows the most dominant effects followed by His and then by Lys. Fragment ions containing basic residues produce more intense peaks than those without basic residues. Doubly charged precursor ions have generally been modeled as producing only singly charged fragment ions, but fragment ions that contain two basic residues may accept both protons during fragmentation. By characterizing the influence of basic residues on gas-phase fragmentation of peptides, this research makes possible more accurate fragmentation models for peptide identification algorithms.

Trypsin digestion is a standard step in many proteomic experiments. Proteins digested by this enzyme are predominantly cleaved C-terminal to arginine and lysine residues, yielding predictable collections of peptides.1 Tryptic peptides are particularly favored for tandem mass spectrometry because they yield spectra with prominent y series ions. Database search algorithms such as SEQUEST[2] have been tuned to work best with tryptic peptides. Alternative digestion techniques can be useful to analyze proteins that do not have many suitable cleavage sites for trypsin (such as membrane proteins) or to increase the diversity of peptides covering a protein region of interest. The cleavage specificities of other digestion enzymes may yield peptides that fragment differently than those produced by trypsin.

One technique showing promise for proteomics is proteinase K digestion.[3] Proteinase K's specificity is more variable than trypsin, favoring amino acids with aromatic side chains as cleavage sites but targeting other residues as well.[4] The digestion time required is less than

*Corresponding author: The Scripps Research Institute; (phone) 858 784-8876; (fax) 858 784-8883; jyates@scripps.edu.
[†]Department of Genome Sciences, University of Washington, Seattle, WA 98195. Current address: Life Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831-6164.
[‡]Department of Chemistry, University of Arizona, Tucson, AZ 85721-0041.
[§]Diversa, 4955 Directors Place, San Diego, CA 92121.

that of trypsin. Because of proteinase K's broader specificity, the diversity of peptides yielded by these digests is higher than in tryptic digests.

The most common type of fragmentation used for proteomics is low-energy collision-induced dissociation (CID). In this technique, peptide ions of a particular mass-to-charge ratio are isolated and collided with noble gas atoms. Collisions add energy to the peptide ions and they fragment. The cleavage of peptide bonds in this process is generally mediated by the mobility of a proton added during the ionization of the peptides, as explained by the mobile proton model.[5] Once a peptide bond has cleaved, the resulting fragments may briefly form a proton-bound dimer, competing to retain the charge.6 Amino acids with high proton affinities thus play a significant role in fragmentation.

Arginine's side chain is the most basic of all amino acids. Its three basic nitrogens are more prone to sequestering or solvating protons than the side chains of other amino acids. Histidine and lysine are the other two most basic amino acids, but studies have varied in their conclusions as to which is more basic than the other in the gas phase.[7] Most peptides contain one other basic site: the N-terminal amine. Trypsin-produced peptides usually contain a C-terminal base (arginine or lysine), an N-terminal amine, and possibly internal histidine residues (or skipped arginine and lysine residues when digestion is incomplete). When trypsin is not used, greater variability in basic residue content is possible.

Peptide identification algorithms based on SEQUEST rely on fragmentation models to identify the peptide corresponding to a particular spectrum. These models generate theoretical spectra for candidate sequences. The observed spectrum is compared to theoretical spectra to determine the best match. If a fragmentation model is accurate, theoretical spectra look more like their experimental counterparts;[8] an inaccurate fragmentation model yields theoretical spectra that are less useful for discriminating between true and false matches. Because trypsin is the most commonly used digestion enzyme, fragmentation models generally create theoretical spectra by rules for tryptic peptides.[2] For example, doubly charged precursor ions are sometimes modeled as producing only singly charged fragment ions. When these algorithms are used on nontryptic peptides, their ability to discriminate between true and false identifications is reduced by their reliance on tryptic assumptions.[9]

Generalizations regarding the effects of basic residues on fragmentation have largely been formed from limited sets of defined peptides[5,10] or from databases of tryptic peptides;[11,12] a systematic study of a large database of peptides with nontryptic basic residue positions has not been conducted. In this report, we statistically characterize the influence of basic residue position and identity on fragment ion intensities in tandem mass spectra. Protein lysates from a variety of organisms were digested by proteinase K digestion. The best identifications from these digests were grouped by their basic residue content, and the relationship between basic residue position and fragmentation was characterized. Arginine, histidine, and lysine show effects of different magnitude and reproducibility. We show that the fragment ions containing basic residues generally produce more intense peaks than fragment ions lacking basic residues. We also note that peaks for doubly charged ions are most intense for fragments containing two basic residues. By statistically analyzing these phenomena, we hope to create a foundation upon which improved fragmentation models can be based.

EXPERIMENTAL SECTION

To improve the statistical power of this analysis, we sought to gather as diverse a collection of peptides as possible. At the same time, we endeavored to limit the size of the spectral collection to keep a high ratio of correct identifications to total spectra included. We achieved this by pooling peptides from a diverse collection of samples, all digested with proteinase K and analyzed by two-dimensional liquid chromatography/ion trap tandem mass spectrometry.

Sample Selection

Chlamydia trachomatis reticulate body whole cell lysates13 were digested by proteinase K and analyzed in five separate 12-cycle MudPITs[14] on a Thermo Finnigan LCQ Deca (San Jose, CA). The results of these five separations were compared by the Contrast algorithm under low cutoffs (described below), revealing that one run identified a larger number of proteins than the others. For this MudPIT, 782 of 33 546 spectra passed the high cutoffs (described below) and were incorporated into the spectral collection for this study (see Table 1 for a summary of the samples included in the study).

The malaria mosquito *Anopheles gambiae* midgut proteome was digested by proteinase K and analyzed by a single 12-cycle MudPIT.[15] Multiple runs were not conducted, but this analysis produced a rich enough collection of peptide identifications for inclusion. Of the 35 516 spectra, 531 yielded identifications that exceeded the high cutoffs.

The brains of male Sprague–Dawley rats were dissected, and the hippocampal proteins were digested by proteinase K.[3] Five separate 12-cycle MudPITs on an Thermo Finnigan LCQ Deca were used to identify the peptides in this sample. The best analysis produced 86 846 spectra, yielding 1488 identifications above the high cutoffs.

Informatics

The spectra produced by each of these experiments were processed through a pipeline of algorithms. The possible charge states of multiply charged peptides were corrected by the 2to3 algorithm.[16] Spectra were matched to peptide sequences by the normalized SEQUEST algorithm,[17] a variant of the software that produces XCorr scores that do not show sequence-length dependence. In evaluating spectra for inclusion, two different sets of criteria were used in the DTASelect and Contrast algorithms.[18] The “low” set used the settings −1 0.3 −2 0.3 −3 0.3 −Smn 5. This setting imposes a moderate cutoff normalized XCorr of 0.3 for peptide inclusion, limiting the listed peptides to sequences containing at least five residues and accepting the default of two peptides per protein for listed proteins. The “high” set, −Smn 5 −a false −o −c 2 −C 2 −2 0.35 −p 1, used the same sequence length cutoff of five residues, but these critieria include only peptides from doubly charged precursor spectra scoring at least 0.35 in normalized XCorr, listing peptides even if they were the only peptide sequences present from their proteins. The low criteria were employed solely for sample assessment while the high criteria were used to select sets of identifications for statistical analysis.

The spectra passing the high criteria were separated into subsets based on their basic residue content. Peptides that contained no basic residues comprised the first group. The next three subsets contained spectra for peptides with a single basic residue; those containining arginine were separated from those containing histidine and from those containing lysine. Six subsets of spectra for peptides containing two basic residues were formed. The first three subsets included identifications that contained two different basic residues, and the other three were identifications containing two of the same basic residues. The spectra in each subset were analyzed by a modification of the DaughterDB algorithm.[11] The measures reported for each subset were processed by scripts created in the *R* statistical environment.[19]

RESULTS AND DISCUSSION

We assembled a set of 155 908 tandem mass spectral identifications from proteomic experiments employing proteinase K (see Table 1). The tandem mass spectra were matched to sequences using the normalized SEQUEST algorithm, which works best when a large percentage of b and y ions can be found in the spectrum. DTASelect isolated sets of identifications from this collection that totalled 2568 peptides. When multiple spectra for a

particular sequence were observed, the highest scoring single spectrum was selected for analysis. Each identification was required to be from a doubly charged precursor and to have a normalized XCorr of 0.35 or higher. This aggressive filtering was required to ensure an acceptable proportion of the identifications were correct.

Most proteomic data sets contain a large number of spectra that cannot be identified successfully.[20] Proteinase K digests are even more problematic, in part because nonspecific digestions yield a greater diversity of peptides, leading to lower concentrations of particular peptide ions for isolation and fragmentation by the mass analyzer. In addition, database identification algorithms may discriminate true and false matches less successfully for nontryptic peptides. Although the criteria used in selecting peptides for this analysis retained fewer than 2% of the raw identifications, more lenient criteria would dilute the results with increased numbers of false matches. The included identifications received XCorr scores that indicate large numbers of fragment ions were matched successfully between theoretical and observed spectra, suggesting that these spectra resulted from mobile proton fragmentation processes rather than charge-remote mechanisms.

Basic Residue Position and Series Intensity

We began our investigation by comparing the intensities of b and y ions with respect to the position of the basic residue. Tryptic peptides generally have basic residues at the C-termini, and they are known to display prominent y ions. Because we used proteinase K, the basic residues among our peptides were positioned more randomly within the sequences. Sequences with N-terminal basic residues were defined to have basic residue positions of 0.0, while those with C-terminal basic residues were assigned positions of 1.0. Basic residues that fell in the middle of the sequences had intermediate positions. The percentage of the spectral intensity for each spectrum corresponding to +1 charged b series fragment ions was stored in the variable *b*, while the percentage corresponding to y ions was stored in variable *y*. The y series enhancement score was calculated to be

$$(y - b)/(y+b)$$

By this expression, a spectrum containing y ions and no b ions would score 1.0, while a spectrum consisting of b ions and excluding y ions would score −1.0. Spectra with equally intense b and y ions would score 0.0. Figure 1 shows the plots of y enhancement versus basic residue position for peptides containing single basic residues.

Linear regressions and correlations characterized the relationship between basic residue position and ion series intensity (see Table 2). In this analysis, the slope of the modeled line can be interpreted as the magnitude of the relationship between basic position and ion series intensity, while the correlation value describes the reproducibility with which these effects appear. Peptides containing arginine showed both the strongest correlation ($R = 0.754$) and highest slope ($m = 0.846$) between the position of the residue and the intensities of the b and y series. Histidine trailed behind arginine with a weaker correlation ($R = 0.704$) and slope ($m = 0.775$). Lysine's position showed a much milder influence on series intensity, with a lower correlation ($R = 0.541$) and slope ($m = 0.490$). Although lysine and histidine may be quite similar in proton affinity,[7] histidine appears to play a more significant role in augmenting fragment ion series.

Subsets of these spectra were examined to determine typical ion series augmentation for the extreme cases when basic residues could be found at one of the termini of the peptides (see Table 2, columns six and seven). When the basic residues were at the N-termini of the peptides, the b ion series was more intense than the y ion series, as indicated by the negative y series

enhancement score. When the basic residues were at the C-termini of the peptides (as in most tryptic peptides), the y series dominated, regardless of the identity of the basic residue. These amino acids appear to play important roles in determining which ion series dominates tandem mass spectra.

Competitive Enhancement

As a further test of the influence of basic residues in fragmentation, we examined sets of peptides that contained two different basic residues (see Table 3). Three combinations of basic residues were examined: His with Lys, His with Arg, and Lys with Arg. Each set of spectra was analyzed twice, tracking each of the two basic residues present. Again, linear regressions were produced to relate the sequence position of the tracked residue to the intensity of the fragment ions produced. This analysis attempted to determine how pairs of basic residues found in a single peptide competed in intensifying fragment ion series.

In both sets of spectra for peptides containing arginine residues, the position of this residue dominated the other in determining which ion series was more intense. In fact, arginine's sequence position correlated with series intensity even more when other basic residues were present than it did when arginine was the only basic residue in the peptide (compare Table 3 to Table 2). The slopes of the linear regressions were also higher, indicating that series intensities reflected the location of arginine more prominently when additional basic residues were present.

The remaining set compared the influences on fragmentation of histidine and lysine when these two basic residues were found in the same peptide. As suggested by the results for peptides containing single basic residues, histidine's influence proved stronger than lysine's; the correlation and slope of its linear regression were both higher. The values reveal, though, that lysine can compete more effectively with histidine than either can with arginine; the slope and correlation calculated for histidine in peptides containing both histidine and lysine are both much lower than those in peptides containing only histidine.

Fragment Ion Augmentation by Basic Content

An amino acid residue near the N-terminus will be found in a higher proportion of b ions than of y ions, while residues near the C-terminus are found in more y than b ions (see Figure 2). We hypothesized that the intensity enhancement of one ion series over another was the result of the augmentation of fragment ions containing the basic residues rather than an increase in intensities of the entire series of ions. We analyzed several sets of spectra in an effort to evaluate this claim: peptides containing no basic residues, peptides containing single basic residues, peptides containing two different basic residues, and peptides containing two of the same basic residues. Of the 2788 identifications passing the high criteria, 2568 were accounted for by these classifications. For each set, we separated the fragment ions by charge state and basic content; singly charged fragment ions were treated separately from doubly charged ones, and within these classifications, fragment ions containing zero basic residues were separated from those containing one basic residue and from those containing two basic residues. Fragment ions from the b and y series were treated interchangeably for the analysis. The average intensity for fragment ions in each class was expressed as a percentage of spectral intensity (see Table 4).

Table 4 reveals several important trends. For sequences containing any basic residues, the singly charged fragment ions containing the basic residues produced more intense peaks than those that lacked them. In sequences that contained only a single arginine, histidine, or lysine, doubly charged fragment ions that incorporated the basic residue produced more intense peaks as well. When peptides contained two basic residues, the peaks for singly charged fragment ions containing both of these residues were not appreciably more intense than fragment ions

containing only one of them. Fragment ions that contained two basic residues showed substantially more intense doubly charged fragment ion peaks than those containing zero or one basic residue. Interestingly, fragment ions containing two arginine residues produced more intense doubly charged fragment ion peaks than singly charged ones, suggesting that when both arginines are located to one side of the cleaved peptide bond, both protons typically associate with the fragment containing the arginines.

Doubly charged precursor ions are often modeled to produce pairs of singly charged fragment ions when they fragment. Until recently, the SEQUEST algorithm generated only singly charged fragment ions in theoretical spectra for doubly charged precursors. Table 4 suggests that doubly charged fragment ions may be prominent in spectra for doubly charged precursor ions containing multiple basic residues. Although these results profile identifications from proteinase K digestions, peptides from trypsin digestions may also contain multiple basic residues; internal histidine residues may be found in many peptides, and incomplete cleavage can result in internal lysines and arginines as well. Doubly charged fragment ions may be significant in spectra from doubly charged precursors.

Chemical Explanation

Basic residues can affect fragmentation in a variety of ways. Their presence in a peptide raises the minimum energy requirement for fragmentation.[5] Their side chains may be the sites in a peptide to which protons show the greatest affinity,[7] in some cases resulting in proton sequestration and enabling charge-remote fragmentation mechanisms.[21–23] Basic side chains may also form salt bridges[24] or charge solvation structures[25] within the peptide, facilitating the transfer of protons to other portions of the peptide. The protonated side chains of histidine and arginine may also cleave adjacent amide bonds to form b ions of non-oxazolone structure.[26,27] As a result, the locations and types of basic residues are pertinent to the fragmentation of peptides.

When a peptide fragments during collision-induced dissociation, proton affinities of the peptide's side chains determine the most stable locations for the proton or protons.[27,28] When a peptide bond cleaves, the resulting fragments form a protonbound dimer,[6] and the fragment containing the most basic site is most likely to retain the charge. If only one proton is carried by the peptide, this process will determine which ion is observed, the b or the y, while the other fragment remains neutral. Thus, the presence and position of basic residues can alter the relative intensities of the b and y ion series. Proton affinity enables Arg to show the most pronounced effect in peptides where two different basic residues are present (see Table 3).

In the case of multiply charged precursor ions, the protons may repel each other, dispersing them to both fragments. This effect favors the coproduction of complementary ions for each cleavage when both fragments are of sufficient mass and basicity to accommodate protons. In some cases, Coulombic repulsion may influence the effective basicity of a fragment. The cleavage of the doubly charged peptide AEGLIRR at its fifth peptide bond, for example, is more likely to produce a b5/y2 pair of singly charged fragment ions than a doubly charged y2 ion, despite the fact that the y2 contains both basic residues in this peptide. Because the arginine residues are adjacent, simultaneous protonation of both is less likely. In peptides where a pair of arginines is separated by other residues, simultaneous protonation becomes more favorable. Though Table 4 indicates that fragments containing two arginines tend to attract both protons more often than not, the proximity of these residues must be taken into account to predict a specific peptide's fragmentation accurately.

The fragmentation influences of the basic residues are a result of several properties. The existence of resonance structures for protonated arginine and histidine allows these residues to distribute the proton's charge for greater stability. Protonated arginine can be shown in four

different resonance structures, while three different structures can be constructed for protonated histidine. Protonated lysine, on the other hand, does not have these resonance structures. All three basic amino acids actively form hydrogen bonds, causing the peptides containing them to adopt secondary structure. Basic side chains may intramolecularly transfer protons to less basic groups, potentially inducing charge-directed fragmentation at the transfer site. Multiple properties of these residues combine to yield their influences on fragmentation.

CONCLUSIONS

The ion trap fragmentation spectra of tryptic peptides contain prominent y series ions because they contain C-terminal basic residues. When basic residues are dispersed throughout the peptide sequence, the fragment ions containing basic residues are generally more intense than the ions lacking them. As a result, peptides with N-terminal basic residues will show intense b ions rather than the intense y ions typical for tryptic peptides. If a peptide contains two different basic residues, one basic residue generally dominates in determining the intensities of fragment ions. Arginine is the most dominant basic residue in this respect, followed by histidine and then lysine. Fragment ions that contain two basic residues may adopt a double charge even when the precursor ion is doubly charged.

These results should be considered in the context of the process that produced them; for example, ion trap collision-induced dissociation excites peptides in a different way than does triple-quadrupole CID. Since fragment ions may be excited along with precursor ions in multiple collision triple-quadrupole CID, one might expect some differences between the spectra produced by these two instruments. The time scale over which fragment ions form and are scanned in an ion trap may also play a role in determining which ions are observed. The role of basic residues as characterized by this article, then, should be taken to represent ion trap fragmentation and may vary in other instruments.

Peptide identification algorithms such as SEQUEST typically generate both b and y fragment ions for each peptide bond. The modeled intensities for these fragment ions are constant for each series; y ions are modeled at a characteristic intensity, and b ions are given a separate, uniform intensity. While this is generally adequate for identifying peptides from tryptic digestions, peptides with more varied basic residue positions may not be identified as successfully with this model. By generating more accurate theoretical spectra for sequences, SEQUEST could identify nontryptic peptides more accurately,[8] separating correct sequences from incorrect sequences by larger margins.

One simple way to improve the SEQUEST fragmentation model would be to use a set of intensities for fragment ions, with the appropriate intensity for each fragment chosen from this set based on the basic residue content of the fragment ion. For example, the model could use a low intensity for singly charged fragments lacking any basic residues and a higher intensity for singly charged fragments that contain one basic residue. Fragment ions that contain multiple basic residues could be modeled in their doubly charged forms as well. In this way, more accurate theoretical spectra for both tryptic and nontryptic peptides could be generated.

A more powerful way to improve these spectrum models would simulate peptide fragmentation. Initially, the software would identify which peptide bonds would be most likely to cleave based on sequence context.[11,12] Given breakage at a particular peptide bond, the model would assess the probabilities of each proton departing with the N-terminal fragment (to produce b and associated fragment ions) or with the C-terminal fragment (to produce y ions). To simulate fragmentation in this way would take more processing power than existing fragmentation models, but it has increased potential to create accurate theoretical spectra and to yield new information about the processes at work in peptide fragmentation.

While proton affinities of the basic amino acids cannot be abstracted directly from the results of this research, it is clear that proton affinity plays a role in determining the intensities of fragment ions. This analysis of database-identified peptides affirms arginine's unique basicity and adds to the discussion of the relative basicities of histidine and lysine. The union of theoretical and empirical studies will help to clarify the role of basicity in peptide fragmentation.

Acknowledgments

We thank Christine C. Wu for providing spectra from *C. trachomatis*. Thanks go to Christine C. Wu and Michael J. MacCoss for generating the rat hippocampal lysate spectra. Laurence Florens graciously provided spectra from the mosquito midgut. D.L.T. was supported by NIH Grant R33 CA81665. Y.H. was supported by NIH GM R0151387 to V.H.W. and by the 2003 Pfizer Graduate Research Fellowship in Analytical Chemistry. Funding was also provided by NIH Grants RR11823-08 and R01MH067880.

References

1. http://www.abrf.org/ResearchGroups/MassSpectrometry/EPosters/ms97quiz/SequencingTutorial.html
2. Eng JK, McCormack AL, Yates JR III. J. Am. Soc. Mass Spectrom 1994;5:976–989.
3. Wu CC, MacCoss MJ, Howell KE, Yates JR 3rd. Nat. Biotechnol 2003;21:532–538. [PubMed: 12692561]
4. Kolattukudy, PE.; Sirakova, TD. Handbook of Proteolytic Enzymes. Barret, AJ.; Rawlings, ND.; Woessner, editors. San Diego: Academic Press; 1998. p. 322-325.
5. Dongre AR, Jones JL, Somogyi A, Wysocki VH. J. Am. Chem. Soc 1996;118(8):8365–8374.
6. Paizs B, Suhai S. Rapid Commum. Mass Spectrom 2002;16:1699–1702.
7. Afonso C, Modeste F, Breton P, Fournier F, Tabet J-C. Eur. J. Mass Spectrom 2000;6:443–449.
8. Huang, Y.; Triscari, JM.; Wysocki, VH. Proc. 50th Am. Soc. Mass Spectrom. Orlando, FL: 2002. p. TPE-135
9. Tabb DL, Saraf A, Yates JR III. Anal. Chem 2003;75:6415–6421. [PubMed: 14640709]
10. van Dongen WD, Ruijters HFM, Luinge H-J, Heerma W, Haverkamp J. J. Mass Spectrom 1996;31:1156–1162. [PubMed: 8916424]
11. Tabb DL, Smith LL, Breci LA, Wysocki VH, Lin D, Yates JR III. Anal. Chem 2003;75:1155–1163. [PubMed: 12641236]
12. Kapp EA, Schütz F, Reid GE, Eddes JS, Moritz RL, O'Hair RAJ, Speed TP, Simpson RJ. Anal. Chem 2003;75:6251–6264. [PubMed: 14616009]
13. Caldwell HD, Kromhout J, Schachter J. Infect. Immun 1981;31:1161–1176. [PubMed: 7228399]
14. Washburn MP, Wolters D, Yates JR III. Nat. Biotechnol 2001;19:242–247. [PubMed: 11231557]
15. Sinden RE, Butcher GA, Beetsma AL. Methods Mol. Med 2002;72:25–40. [PubMed: 12125122]
16. Sadygov RG, Eng J, Durr E, Saraf A, McDonald H, MacCoss MJ, Yates JR III. J. Proteome Res 2002;1:211–215. [PubMed: 12645897]
17. MacCoss MJ, Wu CC, Yates JR III. Anal. Chem 2002;74:5593–5599. [PubMed: 12433093]
18. Tabb DL, McDonald WH, Yates JR III. J. Proteome Res 2002;1:21–26. [PubMed: 12643522]
19. Hornik K. The R FAQ. http://www.ci.tuwien.ac.at/~hornik/R/ISBN:3-901167-51-X.
20. Simpson RJ, Connolly LM, Eddes JS, Pereira JJ, Moritz RL, Reid GE. Electrophoresis 2000;21:1707–1732. [PubMed: 10870958]
21. Tsaprailis G, Nair H, Somogyi A, Wysocki VH, Zhong WQ, Futrell JH, Summerfield SG, Gaskel SJ. J. Am. Chem. Soc 1999;121:5142–5154.
22. Gu CG, Tsaprailis G, Breci L, Wysocki VH. Anal. Chem 2000;72:5804–5813. [PubMed: 11128940]
23. Huang YY, Wysocki VH, Tabb DL, Yates, J. R. III. Int. J. Mass Spectrom 2002;219:233–244.
24. Vachet RW, Asam MR, Glish GL. J. Am. Chem. Soc 1996;118:6252–6256.

25. Wyttenbach T, Paizs B, Barran P, Breci L, Liu D, Suhai S, Wysocki VH, Bowers MT. J. Am. Chem. Soc 2003;125:13768–13775. [PubMed: 14599216]
26. Farrugia JM, O'Hair RAJ, Reid GE. Int. J. Mass Spectrom 2001;210/211:71–87.
27. Wysocki VH, Tsaprailis G, Smith LL, Breci LA. J. Mass Spectrom 2000;35:1399–1406. [PubMed: 11180630]
28. Polce MJ, Ren D, Wesdemiotis C. J. Mass Spectrom 2000;35:1391–1398. [PubMed: 11180629]

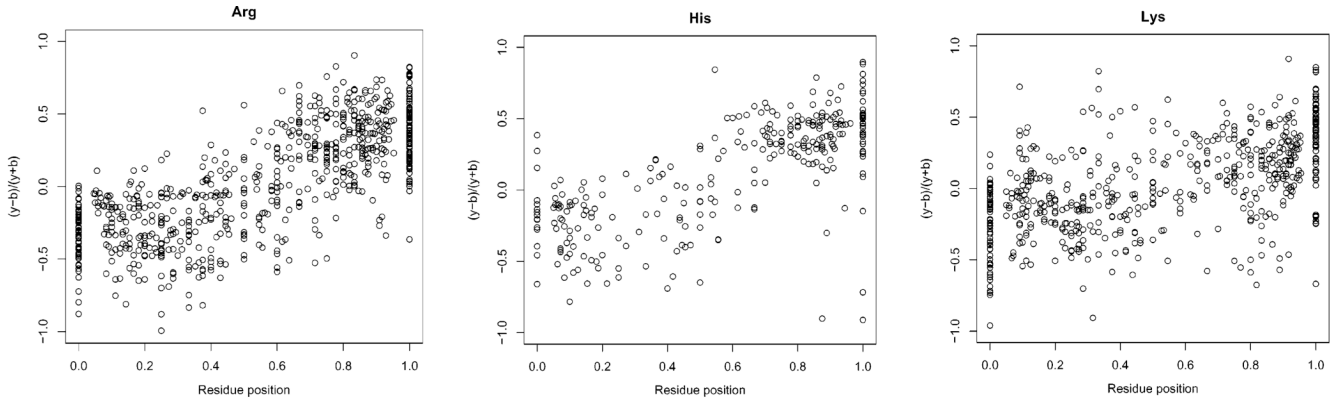

Figure 1.

The three basic residues each showing a relationship between residue position and the enhancement of one fragment ion series over the other. When basic residues are at the N-terminus (left edges of graphs), the b series is most intense. When the basic residues are at the C-terminus (right edges of graphs), the y series is most intense. Arginine is the most clear-cut in its effect. See Table 2 for statistics characterizing these figures.

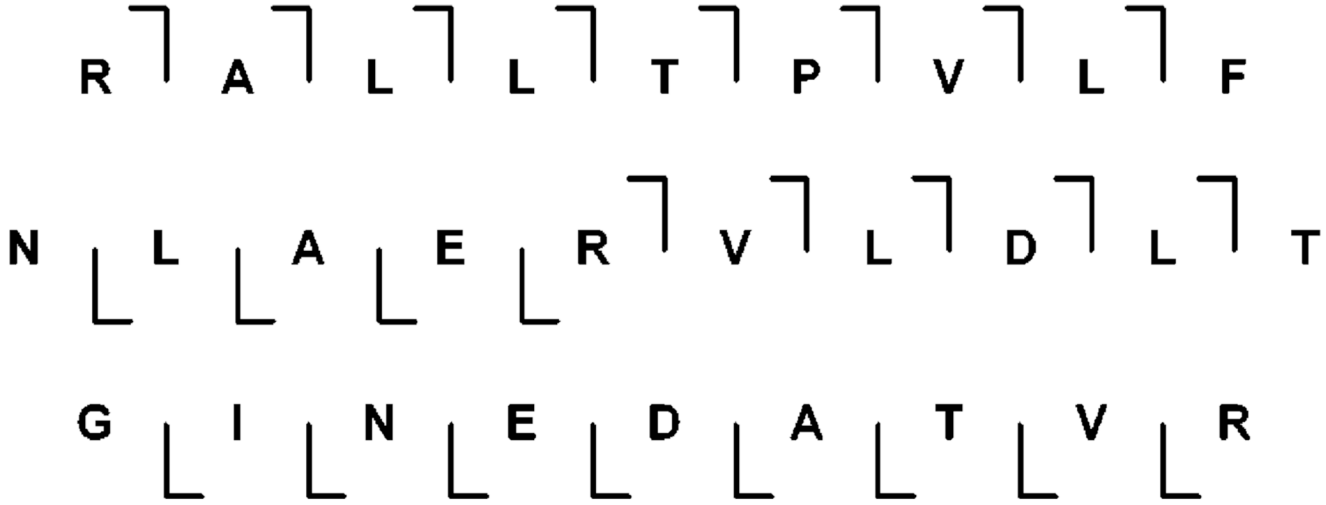

Figure 2.
Each residue contained in half of fragment ions. When the basic residue is at the N-terminus as in the first sequence shown, only the b ions contain the base. Likewise, when the basic residue is at the C-terminus as in the third sequence, it will be found only in y ions. If a basic residue is in the middle of the sequence, a mix of b and y ions will incorporate the base. Either the b or y ion formed at a particular peptide bond contains any particular residue, so a specific amino acid from a peptide is found in half of the possible fragment ions from these two series.

Table 1

Three Proteinase K Digests Compared[a]

sample	chlamydia	mosquito	rat
total spectra	33546	35516	86846
protein IDs, low cutoff	323	132	890
peptide IDs, low cutoff	1282	1109	2827
peptide IDs, high cutoff	782	531	1488

[a]Three sets of spectra were combined to generate the subsets of spectra for examining basic residue content. All three samples were digested with proteinase K and subjected to 12-cycle MudPIT analysis. The rat hippocampus separation yielded the largest number of spectra and identifications. Taken together, 2788 peptides passed the high cutoffs (this is less than the sum of the three because some peptides, such as those for keratin, were observed in multiple samples).

Table 2

Basic Residue Positions Corresponding to Relative Ion Series Intensity[a]

					$(y-b)/(y+b)$	
contain	**exclude**	**count**	**correlation**	**slope**	**N-term**	**C-term**
R	H, K	731	0.754	0.846	−0.347	0.381
H	K, R	285	0.704	0.775	−0.156	0.393
K	H, R	579	0.541	0.490	−0.264	0.354

[a]The first three columns denote which basic residue was contained in the sequences and how many spectra were found matching this pattern. The “correlation” and “slope” columns were calculated from linear regressions of the data found in Figure 1. The next two columns show the mean y series enhancement values calculated for peptides with basic residues at the N-terminus or at the C-terminus.

Table 3

Competition of Basic Residues When Two Are Present[a]

tracking	contain	exclude	count	correlation	slope
R	H, R	K	180	0.826	0.920
H	H, R	K	180	−0.098	−0.105
R	K, R	H	275	0.873	1.171
K	K, R	H	275	−0.171	−0.252
H	H, K	R	98	0.462	0.456
K	H, K	R	98	0.256	0.266

[a]Each set was analyzed twice, once for each base. The “tracking” column indicates which of the two basic residues' positions was used for the linear regression in each row. When arginine was present, its effect was far more pronounced than the other basic residue (as evidenced by arginine’s higher slopes and correlations in the first and second sets). Histidine and lysine positions are positively correlated with series intensity when they are the only basic residues in the peptide (see Table 2), but the presence of arginine changes this correspondence to a negative correlation. When histidine and lysine are the only basic residues in a peptide, histidine produces the higher slope and correlation.

Table 4

Singly and Doubly Charged Fragment Ion Peak Intensities for Each Set of Spectra[a]

fragment sequences			singly charged fragments av intens (%) when containing			doubly charged fragments av intens (%) when containing		
contain	**exclude**	**count**	**0 bases**	**1 base**	**2 bases**	**0 bases**	**1 base**	**2 bases**
	HKR	196	0.80			0.27		
1R	HK	731	0.63	1.09		0.10	0.57	
1H	KR	285	0.57	0.99		0.11	0.46	
1K	HR	579	0.68	0.98		0.10	0.40	
1H, 1R	K	180	0.41	0.79	0.80	0.06	0.11	1.07
1K, 1R	H	275	0.43	0.81	0.90	0.10	0.13	0.82
1H, 1K	R	98	0.46	0.87	0.76	0.04	0.12	0.87
2R	HK	57	0.34	0.95	0.21	0.06	0.15	0.91
2H	KR	35	0.38	0.68	0.80	0.10	0.08	0.73
2K	HR	132	0.56	0.79	0.79	0.07	0.11	0.65

[a]The spectra were divided into those containing no basic residues, those with a single basic residue, and those with two basic residues. No distinction was made between b and y fragment ions. The average fragment ion intensity in each of these categories was calculated and expressed as a percentage of the spectrum's intensity. For reference, a typical y ion in a tryptic spectrum accounts for ~1% of spectral intensity while a typical b ion accounts for approximately half that. DaughterDB does not filter out low-intensity peaks or account for the isotopic variants of fragment ions, and so the reported peak intensities may seem lower than expected. The top row of data show the result for peptides with no basic residues. The spectra from Table 2 are shown in the next section, and spectra from Table 3 are shown next. Finally, spectra for peptides containing two of a particular basic residue are listed. Singly charged fragment ions containing at least one basic residue are more intense than those without any basic residues. Doubly charged fragment ions can be as intense as singly charged fragment ions when two basic residues are contained in the fragment ion sequence.